

Centro Servizi Formativi di Crema

Area Tecnologie Informatiche



Anno Formativo 2002/2003

Corso post-diploma per Tecnico di Reti Webmaster

Internet e il Web

a cura di Sergio Fumich

Indice

1 Internet e il World Wide Web

Oggi si accede a Internet per lo più utilizzando un'applicazione chiamata *browser Web*. I browser Web più comuni sono Internet Explorer e Netscape Communicator.

Il browser, che consente di visualizzare le pagine Web su Internet, è un esempio di *applicazione client*. Le applicazioni client vengono eseguite sul computer locale e consentono al proprio computer di comunicare con altri computer collegati in Internet.

Molti computer in Internet sono dei *server* che forniscono programmi ai client.

Alcuni esempi di programmi basati su server sono quelli di posta elettronica e di pubblicazione di pagine Web.

I client di posta elettronica come Outlook Express, consentono di accedere alla propria posta memorizzata sul server di posta elettronica. I browser Web consentono di accedere alle pagine Web memorizzate su un server Web.

La programmazione che consente alle applicazioni client di comunicare con le applicazioni server attraverso una rete viene indicata col termine *programmazione client-server*. Oggi Internet è una rete di reti e fornisce molti tipi di programmi client-server.

2 Storia di Internet

Internet, nella sua forma originaria ARPANET, era costituita da un numero relativamente piccolo di computer che collegavano fra loro le agenzie governative statunitensi ed alcune università. La rete veniva utilizzata principalmente per la ricerca scientifica e l'educazione.

Nel settembre 1969 ARPANET era costituita da un'unica rete tra UCLA, UCSB, Utah e lo Stanford Research Institute, ed era fondamentalmente una rete *host-to-host*.

Ciascun computer veniva indicato con il termine *host* e disponeva di una propria serie di programmi. I singoli utenti dovevano collegarsi ad esso per accedere ai programmi. Non esistevano client in questo schema di rete.

Nel 1971, negli Stati Uniti, ARPANET comprendeva meno di due dozzine di università. Nel 1976 Internet venne aperta al traffico commerciale. A quel tempo la rete veniva principalmente utilizzata per la posta elettronica e Telnet.

I programmi Telnet consentono di accedere per via remota a un computer attraverso Internet. Gli studenti di molte università utilizzavano comunemente Telnet per accedere alla rete attraverso i terminali locali. Più tardi, il protocollo FTP (*File Transfer Protocol*) diventò l'applicazione utilizzata comunemente.

Un server FTP consente ai client FTP di trasferire i file da e verso il server. FTP fa parte di una suite di protocolli chiamata TCP/IP (*Transmission Control Protocol / Internet Protocol*).

Alla fine degli anni '80 il numero dei computer host presenti in Internet aumentò fino a raggiungere 100.000 unità. TCP/IP divenne lo standard di fatto per i protocolli Internet. I nuovi programmi Internet facevano affidamento su TCP/IP per instradare e distribuire le informazioni. Risolvendo il problema di come distribuire le informazioni attraverso Internet, TCP/IP consentì ai programmatori di creare un numero maggiore di applicazioni.

Internet si sviluppò fino a diventare la più grande rete di reti (*internetworking*) con l'avvento dei fornitori di accesso a Internet (ISP o *Internet Service Provider*). Gli ISP fornivano il software client che consentiva al singolo utente di collegarsi alla rete anche se non era uno studente o un impiegato governativo.

Il 12 giugno 1991, una nuova tecnologia fornì uno strumento che consentiva di localizzare e accedere facilmente ai documenti presenti su varie piattaforme attraverso Internet. Questi documenti che abbracciavano piattaforme differenti, vennero chiamati *documenti ipertestuali*, e potevano essere consultati indipendentemente dalla piattaforma informatica utilizzata.

Questa tecnologia coinvolse un nuovo linguaggio standard, chiamato HTML (*HyperText Markup Language*) ed un nuovo protocollo HTTP (*HyperText Transfer Protocol*). Le istruzioni HTML, chiamate *tag* (o marcatori), indicavano ai browser Web come formattare ed elaborare un documento ipertestuale.

Nel 1993 l'NCSA (*National Center for Supercomputing Applications*) dell'Università dell'Illinois rilasciò NCSA Mosaic, il primo browser Web commerciale.

Attualmente i due browser più utilizzati sono Microsoft Internet Explorer e Netscape Communicator. HTTP è il protocollo di comunicazione utilizzato da Internet per trasferire i documenti ipertestuali. I documenti ipertestuali condivisi su Internet vengono chiamati *pagine Web* e l'indirizzo della locazione di questi documenti è noto con il termine URL (*Uniform Resource Locator*).

Le pagine Web solitamente contengono testi, immagini, elementi multimediali e collegamenti ad altre pagine Web. Attualmente il W3 Consortium definisce e gestisce gli standard HTML e HTTP, così come molti altri standard di Internet.

Oggi Internet è descritta come un sistema di informazioni globale che utilizza TCP/IP per fornire servizi di comunicazione. Le applicazioni Internet vanno ben oltre la posta elettronica e i servizi per la pubblicazione sul Web. Internet viene utilizzata per inviare fax, visualizzare filmati e fornire servizi di invio di messaggi in tempo reale. Alcuni ISP offrono ai propri clienti la possibilità di trasferire il traffico vocale attraverso Internet.

Molte aziende stanno attualmente sviluppando applicazioni basate su Internet per facilitare le transazioni del commercio elettronico. le applicazioni *business-to-consumer* comprendono carte di acquisto, la gestione delle carte di credito e servizi di gestione di gruppi di utenti. Le applicazioni *business-to-business* includono la gestione di ordini e di magazzino.

Aziende quali MCI WorldCom e AT&T, che attualmente possiedono grandi porzioni di Internet, stanno aggiornando l'infrastruttura della rete in modo che possa gestire un traffico sempre maggiore.

Una rappresentazione grafica di tutto il traffico Internet è disponibile presso il sito Internet Weather (http://www.internetweather.com). Questo sito aiuta a identificare le aree in cui il traffico Internet è congestionato.

3 Funzionamento di Internet

I computer collegati in Internet utilizzano hardware, software, programmi e sistemi operativi differenti. Per poter comunicare su Internet, i computer utilizzano una suite di protocolli nota con l'acronimo TCP/IP.

La suite TCP/IP comprende una varietà di altri protocolli, tra cui TCP, IP, Telnet, FTP, HTTP, SMTP (*Simple Mail Transfer Protocol*) e NNTP (*Network News Transfer Protocol*).

Mentre TCP/IP stabilisce le regole di base per comunicare attraverso Internet, altri protocolli di rete, quali NetBEUI e IPX/SPX, stabiliscono le regole di comunicazione tra computer all'interno di reti private (*intranet*). Questi protocolli non consentono ai computer di comunicare attraverso Internet.

Tutti i computer che si collegano ad Internet, sia client sia server, devono utilizzare il protocollo TCP/IP.

Poiché TCP/IP è un protocollo scalabile che può essere utilizzato per le LAN (*Local Area Network*) e le WAN (*Wide Area Network*), la maggior parte dei sistemi operativi ha già incorporato al suo interno il software TCP/IP.

4 TCP

Il protocollo TCP (*Transmission Control Protocol*) controlla la trasmissione fisica dei dati attraverso Internet.

TCP suddivide la trasmissione in piccoli pacchetti, che vengono inviati tramite Internet verso un computer di destinazione. Il software sul computer di destinazione rimette insieme i pacchetti, ricreando in tal modo la trasmissione originale.

Nel percorso verso il computer di destinazione, ciascun pacchetto di dati effettua diverse fermate, chiamate *hop*, presso computer situati lungo il percorso. Il computer che si trova a ciascuna fermata viene indicato come *gateway*, perché traccia la tappa successiva del percorso e invia il pacchetto per la sua strada. Questo processo viene ripetuto per ciascun computer gateway, finché il pacchetto raggiunge la propria destinazione.

5 IP

Il protocollo IP (*Internet Protocol*) utilizza uno schema specifico per gestire l'indirizzamento di ogni singolo pacchetto e per assicurarsi che tutti i pacchetti giungano a destinazione.

I computer gateway utilizzano l'indirizzo IP per stabilire come tracciare la successiva tappa che il pacchetto dovrà percorrere. Di conseguenza, ciascun pacchetto di una trasmissione può intraprendere un percorso differente in Internet, ma alla fine tutti raggiungeranno la stessa locazione e verranno riassemblati nel messaggio originale.

Virtualmente tutto il traffico Internet viaggia attraverso questo modello di instradamento (*routing*) del protocollo IP.

6 Indirizzo IP

Ciascun computer collegato a Internet, compresi i client, ha un *indirizzo IP* associato.

L'indirizzo IP viene utilizzato dai server di routing in Internet per determinare come instradare e distribuire i pacchetti Internet. Ciascun server Web possiede un numero di computer unico.

Gli ISP assegnano un indirizzo IP statico o dinamico ogni volta che un computer client si collega a Internet.

Un indirizzo IP dinamico è un indirizzo IP che cambia a ogni collegamento. Un indirizzo IP statico viene assegnato a un computer specifico e non cambia. La maggior parte degli ISP più importanti assegna indirizzi IP dinamici ai propri client.

Un server Web deve avere un indirizzo statico in modo che i server di routing su Internet siano sempre a conoscenza dell'indirizzo IP che conduce al sito Web.

Il protocollo IP definisce un indirizzo IP come un numero a 32 bit costituito da quattro gruppi di numeri separati da punti. Questi quattro numeri sono compresi tra 0 e 255. L'indirizzo IP rappresenta la versione abbreviata di un numero binario long.

Alcuni indirizzi IP sono riservati per scopi speciali.

Ad esempio, l'indirizzo IP 127.0.0.1, che viene indicato con il termine *localhost*, viene utilizzato per instradare i pacchetti verso il computer locale. È usato spesso per controllare se TCP/IP è stato correttamente installato e per verificare che la scheda dell'interfaccia di rete funzioni.

I programmatori Internet utilizzano in fase di sviluppo l'indirizzo IP localhost per rappresentare il server Web locale. Digitando http://127.0.0.1 o http://localhost nel browser, ci si indirizza verso il server Web sul computer locale.

Il protocollo IP è in fase di aggiornamento verso una nuova versione che sia in grado di gestire un maggior numero di indirizzi IP rispetto l'attuale.

7 Protocollo HTTP

Il protocollo HTTP viene utilizzato per inviare documenti HTML attraverso Internet. Questo protocollo invia i documenti suddividendoli in pacchetti, utilizzando TCP/IP.

A ciascun pacchetto, il protocollo HTTP associa un'intestazione, contenente informazioni quali il nome e la locazione della pagina da richiedere, il nome e l'indirizzo IP del server remoto che contiene la pagina Web, l'indirizzo IP del client locale, il numero della versione HTTP e l'URL della pagina.

Il protocollo HTTP versione 1.0 è un protocollo privo di informazioni sullo stato. Ciò significa che quando un client richiede un documento dal server Web, il server Web fornisce la pagina al client e pone fine a tutte le comunicazioni con quest'ultimo.

Nella versione 1.1 del protocollo HTTP, il server Web e il client sono in grado di gestire la propria connessione attraverso le pagine Web.

8 Programmi di utilità di TCP/IP

Quando il client richiede un servizio da un server su Internet, ad esempio l'invio di una pagina Web, utilizza il modello client/server di TCP/IP.

Esistono molti programmi di utilità in TCP/IP utilizzati per effettuare il monitoraggio della rete. Alcuni di questi programmi servono per eseguire test diagnostici sul proprio sito.

I due programmi di utilità TCP/IP più importanti sono *ping* e *tracert*.

8.1 Ping

L'indirizzo IP del sito Web può essere recuperato utilizzando il programma di utilità TCP/IP chiamato *ping*, che viene utilizzato per eseguire il test del collegamento alla rete.

Quando si effettua il ping di un sito, vengono inviati quattro pacchetti all'indirizzo IP. Se il sito Web è attivo e collegato a Internet, ping restituisce il suo indirizzo IP. Se il sito Web non è collegato, allora compare un messaggio di tipo *Request Timed Out*.

8.2 Tracert

Un altro programma di utilità TCP/IP è *tracert*, utilizzato per localizzare il server Web è quello che consente di ottenere una traccia dei pacchetti.

Questo programma fornisce il percorso che il pacchetto Internet percorre per andare dal client al server. Vediamo come funziona.

Tutte le comunicazioni Internet vengono inviate in piccole porzioni di dati chiamate *pacchetti*. Quando il pacchetto si ferma presso un server gateway, questo stabilisce quale percorso deve intraprendere successivamente. Il programma tracert elenca ciascuna fermata, *hop*, che indica che il pacchetto è stato indirizzato verso un altro server o un router.

L'ultimo hop nell'elenco indica l'indirizzo IP del sito Web.

Quando nell'elenco compare un asterisco (*), significa che il pacchetto ha avuto un ritardo su quel server. Quindi tracert può aiutare ad individuare tempi di attesa o ritardi durante l'istradamento dei pacchetti verso il sito Web.

Per chi non avesse tracert sul proprio Pc, all'indirizzo *http://www.net.cmu.edu/cgi-bin/netops.cgi* Carnegie Mellon consente di eseguire il comando ping o tracert verso host di altre reti.

8.3 Software di rete del client TCP/IP

Perché un computer possa fungere da client su Internet, deve avere un software di rete TCP/IP installato.

Client per reti Microsoft è il software di rete per il client utilizzato dai computer con sistema Windows.

8.4 Programmi di utilità di configurazione IP

Le impostazioni TCP/IP quali l'indirizzo IP del client, vengono assegnate in un sistema Windows utilizzando il programma *Rete* localizzato nel *Pannello di controllo*. Il s.o. Windows fornisce i programmi di utilità necessari per rilevare le impostazioni TCP/IP.

Winipcfg è il programma utilizzato dai s.o. Windows 95 e Windows 98 e ME. *Ipconfig* è quello utilizzato dal s.o. Windows NT.

9 Nomi di dominio

Un gruppo di pagine Web, directory e file tra loro correlati, denominato *sito Web*, viene memorizzato in Internet su un *server Web*.

Un *server Web* è un computer che contiene un software che consente di ospitare le pagine Web su Internet.

Il server Web può ospitare più siti Web sullo stesso computer fisico. Il software del server Web è in grado di differenziare i vari siti Web utilizzando un indirizzo IP e un nome di dominio differenti per ciascuno.

Un sito Web può essere identificato dal suo indirizzo IP o dal suo nome di dominio.

Il *nome di dominio* identifica in modo univoco un unico sito Web in Internet. Tutti i nomi di dominio vengono registrati in un database generale noto come *registro*.

All'inizio degli anni '90, chi voleva registrare un nome di dominio doveva farlo presso InterNIC, che elaborava tutte le registrazioni e gestiva il database del registro. Nel 1992 InterNIC ha trasferitola registrazione dei domini e la responsabilità della gestione del registro ad una società chiamata Network Solutions. Attualmente molte altre aziende, i *registrar*, sono autorizzate a registrare il nome di dominio di un sito Web nel database del registro.

9.1 Registrazione di un nome di dominio.

Per registrare un nome di dominio, basta accedere al sito Web di uno dei *registrar*.

Nel sito si trova un modulo per interrogare il database e stabilire se il nome di dominio scelto è disponibile. Se la risposta è positiva, all'utente viene chiesto di riempire un modulo per registrare il dominio.

Le informazioni minime indispensabili da fornire per la registrazione sono:

- Nome del dominio
- Nome, indirizzo e indirizzo di posta elettronica di chi lo vuole registrare
- Nome, indirizzo e indirizzo di posta elettronica del contatto amministrativo
- Nome, indirizzo e indirizzo di posta elettronica del contatto tecnico
- Nome, indirizzo e indirizzo di posta elettronica del contatto per le fatturazioni
- Nome e indirizzo IP dei server DSN primario e secondario (i server di routing)
- Informazioni di fatturazione per il pagamento online.

Il contatto amministrativo, quello tecnico e quello per la fatturazione all'interno del database avrà un *handle* InterNIC. Questo handle o *NIC handle*, viene utilizzato per identificare un singolo o un'organizzazione all'interno del database.

Utilizzando l'handle InerNIC si semplifica il processo di registrazione. La prima volta che viene registrato un dominio, al proprio nome viene assegnato un handle InterNIC. Nel corso delle successive registrazioni di nomi di dominio, si dovrà specificare soltanto il proprio handle, senza inserire una seconda volta le informazioni di contatto.

10 Server DNS

Poiché un server di rete può ospitare più di un server Web, c'è la necessità di un mezzo per identificare i server Web in rete. Il DNS (*Domain Name Service*) è un servizio di rete che traduce i nomi dei domini in indirizzi IP.

A ciascun server Web viene assegnato un indirizzo IP e un nome di dominio. Quando si registra il proprio dominio, non si ottiene l'indirizzo IP per il sito Web, ma il nome e l'indirizzo IP di un server DNS. Quest'ultimo contiene un elenco di nomi di dominio e di indirizzi IP a essi associati.

L'ISP che ospita il vostro sito Web inserirà nel proprio server DNS il vostro nome di dominio e l'indirizzo IP.

Quando un client richiede una pagina Web, il server di rete del client localizza il server DNS appropriato, che poi fornisce l'indirizzo IP corretto.

10.1 Whois

Per localizzare i server DNS, l'handle InterNIC o le informazioni sul *registrant* (il proprietario legale del dominio) relative ad un sito Web, si può usare il programma di utilità *whois.*

Whois ricerca le informazioni associate a singole persone, ai nomi di dominio e ai server DNS all'interno del registro. È possibile accedere al programma di utilità *whois* all'indirizzo *http://www.netsol.com/cgi-bin/whois/whois.*

All'indirizzo di sopra si possono eseguire ricerche per le estensioni *.com*, *.net*, *.org*, *.edu*, *.tv*, *.us*, *.name*, *.cc*, *.bz*,*. ws*, *.de*, *.be*, *.info*, *.biz.*

Per ricerche sul registro USA si può andare all'indirizzo *http://www.arin.net/whois/arinwhois.html*, per quello europeo ci si può rivolgere all'indirizzo *http://www.ripe.net/perl/whois*, per quello asiatico e del Pacifico *http://www.apnic.net/*. Per i server DNS militari si può andare all'indirizzo *http://www.nic.mil/dodnic/contents.html*, per quelli del governo degli Stati Uniti all'indirizzo *http://www.nic.gov/cgi-bin/whois.*

11 Il servizio World Wide Web

Un *server di rete* è un'applicazione server di rete che fornisce servizi correlati con la rete quali i servizi di posta elettronica, i database, la messaggistica e i servizi per i file. Il software di rete che memorizza le pagine Web ed elabora le richieste di tali pagine è chiamato *server Web*.

Gli amministratori di rete possono definire sullo stesso server Web più siti Web. Molte aziende forniscono servizi per ospitare diversi siti Web sullo stesso computer. Queste aziende vengono chiamate *fornitori di servizi hosting* (HSP, *Hosting Service Provider*).

L'architettura della propria applicazione Internet dipende dal tipo di applicazione che si sta sviluppando. La programmazione per Internet è indicata con il termine *programmazione client/server* perché l'applicazione client comunica con l'applicazione server.

Un browser invia una richiesta da una macchina client per un documento Internet collocato su un server Web remoto. Il client comunica con il resto di Internet attraverso un server di rete, che in molti casi è quello dell'ISP. Il server di rete del client istrada i pacchetti della richiesta verso il server Web che contiene la pagina in questione. Il server di rete invia i pacchetti di quanto richiesto

al browser del client. Il browser riceve ed elabora la richiesta ed effettua il *rendering* (visualizza) il documento richiesto.

Oggi, questo modello base è diventato più complesso a causa del numero e del tipo di applicazioni presenti in Internet. Le applicazioni Internet non sono più limitate ad un unico server. Le applicazioni Internet per il commercio elettronico, ad esempio, spesso comunicano con server di database che non sono localizzati sullo stesso server di rete fisico del server Web.

12 Server Web

Esistono molti tipi differenti di server Web, i più comuni sono *Apache* e *Internet Information Server*.

Sul s.o. UNIX, Apache è il server Web più comune. Su piattaforme Windows NT, il più comune è Internet Information Server.

Sui computer desktop con s.o. Windows 98, si utilizza *Microsoft Personal Web* per creare un ambiente Web che faccia da test nella fase di costruzione dei propri siti Web.

Il sito Web Netcraft all'indirizzo *http://uptime.netcraft.com/up/graph/* offre un programma di utilità che consente di rilevare il tipo di server Web e il sistema operativo che il sito Web sta utilizzando.

13 Fornitori di servizi di hosting

Dopo aver selezionato un nome di dominio e averlo registrato per mezzo di un registrar InterNIC, serve un fornitore di servizi che ospiti il sito Web.

Gli ISP offrono servizi Internet, che possono includere servizi di dial-up per collegare i client a Internet e la possibilità di ospitare siti Web. Non tutti gli ISP forniscono servizi di alta qualità.

Un fornitore di servizi di hosting (HSP, *Hosting Service Provider*) è un ISP che si è specializzato nell'ospitare siti Web. Alcuni HSP si sono specializzati nel lavorare con programmatori Internet, mentre altri forniscono solo i servizi generali per ospitare un sito Web.

Molte aziende possiedono un proprio server Web, alcune aziende più piccole spesso inizialmente realizzano una loro presenza su Internet attraverso un fornitore di servizi hosting.

Alcuni HSP offrono la possibilità di acquistare un computer server e mantenerlo presso la loro sede. L'HSP gestirà l'apparecchiatura che collega il server a Internet, ma la responsabilità per quanto riguarda la gestione del software del server resta affidata all'azienda.

La persona che gestisce il server Web è il *Webmaster*.

Prima di iniziare a cercare un HSP, è utile stabilire i criteri da utilizzare per fare una valutazione. È consigliabile elencare una serie di caratteristiche considerate essenziali per il successo del proprio sito Web ed ordinarle per priorità. Un elenco dei criteri di valutazione può aiutare a trovare il fornitore più adatto alle proprie necessità.

Una volta definiti i criteri, è opportuno confrontare tra loro almeno una decina di fornitori differenti.

Esistono in Internet molti elenchi di HSP che spesso forniscono una graduatoria di HSP, basata sulle funzionalità che questi sono in grado di gestire o sul riscontro fornito dai rispettivi utenti.

CNET (*http://www.cnet.com*) fornisce un elenco esteso di fornitori di servizi hosting. Altri siti Web similari sono: HostIndex (*http://www.hostindex.com*), TopHosts.Com (*http://www.tophosts.com*), The List (*http://thelist.internet.com*).

14 Creazione di una pagina Web

Le pagine Web vengono create utilizzando il linguaggio HTML. La versione più recente del linguaggio è la 4.01. HTML è un *linguaggio di markup* che utilizza i *tag* o *marcatori* per indicare al browser come formattare o rappresentare la pagina Web.

Le pagine Web possono essere create con un semplice editor di testi, oppure con un *editor per pagine Web*. Gli editor Web forniscono un'interfaccia utente grafica che consente di gestire facilmente il layout, la grafica e le caratteristiche di una pagina Web. La scelta dell'editor Web dipende dal tipo di pagine Web che si intende creare e dal proprio livello di esperienza.

Una volta create le proprie pagine Web tramite uno strumento di sviluppo Web, è necessario pubblicarle su un server Web per poterle visualizzare. Alcuni strumenti di sviluppo Web consentono di pubblicare le pagine direttamente su un server.

Se si utilizza un editor di testi, serve un programma per copiare i file creati sul server Web. Si può anche utilizzare un software FTP per trasferire i file dal computer client al server Web, ad esempio WS_FTP Pro di Ipswitch (*http://www.ipswitch.com*). Esistono molti altri programmi FTP disponibili in Internet, per trovarli si può visitare ad esempio il sito di Tucows all'indirizzo *http://www.tucows.com*.